



# Dynamic modeling of genes controlling cancer stem cell proliferation

Zhong Wang<sup>1</sup>, Jingyuan Liu<sup>2</sup>, Jianxin Wang<sup>1</sup>, Yaquan Wang<sup>2</sup>, Ningtao Wang<sup>2</sup>, Yao Li<sup>3</sup>, Runze Li<sup>1,2</sup> and Rongling Wu<sup>1,2</sup>\*

<sup>1</sup> Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA, USA

<sup>2</sup> Department of Statistics, The Pennsylvania State University, State College, PA, USA

<sup>3</sup> Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

## Edited by:

Lin S. Chen, The University of Chicago, USA

## Reviewed by:

Qianying Liu, University of Chicago, USA

## \*Correspondence:

Rongling Wu, Center for Statistical Genetics, The Pennsylvania State University, Hershey, PA 17033, USA.  
e-mail: rwu@phs.psu.edu

The growing evidence that cancer originates from stem cells (SC) holds a great promise to eliminate this disease by designing specific drug therapies for removing cancer SC. Translation of this knowledge into predictive tests for the clinic is hampered due to the lack of methods to discriminate cancer SC from non-cancer SC. Here, we address this issue by describing a conceptual strategy for identifying the genetic origins of cancer SC. The strategy incorporates a high-dimensional group of differential equations that characterizes the proliferation, differentiation, and reprogramming of cancer SC in a dynamic cellular and molecular system. The deployment of robust mathematical models will help uncover and explain many still unknown aspects of cell behavior, tissue function, and network organization related to the formation and division of cancer SC. The statistical method developed allows biologically meaningful hypotheses about the genetic control mechanisms of carcinogenesis and metastasis to be tested in a quantitative manner.

**Keywords:** differential equation, cancer stem cell, statistical model, genetic architecture

## INTRODUCTION

Human cancers grow from stem cells (SC) in the way that healthy organs do (Dick, 2003; Dontu et al., 2003; Alison, 2005; Brown et al., 2007; Lobo et al., 2007; Baumann et al., 2008; Visvader and Lindeman, 2008). This discovery, or hypothesis, has refreshed our hope to eliminate these fatal diseases. A stem cell is one that, when it divides, produces two unequal daughters. While one remains a stem cell, the other differentiates into a diverse range of specialized cell types. There is emerging evidence that cancer formation may follow a similar mechanism of cell division (Lapidot et al., 1994; Al-Hajj et al., 2003; Singh et al., 2004). In breast and brain tumors, for example, a minority population of cancer SC has the ability to self-renew, whereas the majority of cancer cells have limited or no ability to proliferate. Because cancer SC may drive the growth and spread of the tumor, the prevention of a malignant tumor can be made possible by designing specific drugs that destroy those mother cells (Clarke and Fuller, 2006). However, current cancer-killing therapies are not able to distinguish cancer SC from malignant cells. If only the descendants of the non-stem-cell daughters are removed or destroyed, the tumor will continue to return. For this reason, the primary step toward the effective control and prevention (and even eradication) of cancers is the prospective identification and characterization of cancer SC, followed by the removal of targeted cancer SC by developing specified drug therapies.

The distinction between stem cancer cells and non-stem cancer cells has been a highly challenging issue, but can be made possible through a profound understanding of the differences in their origin, property, and function (Lobo et al., 2007). Modern life sciences, such as molecular biology, genomics, biomedicine,

or oncology, are powerful for understanding the genetic and molecular mechanisms of carcinogenesis and the complexity of its progress and dynamics based on experimental and clinical observations. The past decade has witnessed a vital development in high-dimensional genetic, genomic, and proteomic technologies that have led to the availability of an enormous amount of data (Clarke et al., 2008). For example, high-throughput single nucleotide polymorphism (SNP) analysis can yield hundreds of thousands of polymorphic markers for a single specimen. Gene microarray techniques, such as Affymetrix, allow concurrent measurements on numerous mRNA transcripts. Dimensionality can be even larger in proteomic studies using the ProteinChip System. These data provide invaluable resources to detect and define expression patterns for the phenotypes of complex traits or diseases (Khan et al., 2001; Bhanot et al., 2005) and predict a potential clinical outcome of patients after a particular drug therapy (Ayers et al., 2004; Lopez-Rios et al., 2006; Ganly et al., 2007).

In another different world, tremendous progress has been made in mathematical modeling of complex biological processes and cell differentiating mechanisms (Anderson and Quaranta, 2008; Piotrowska et al., 2008). Mathematical models and numerical simulations can uncover and explain many still unknown aspects of cell behavior, tissue function, and network organization. Models based on fundamental biological and developmental mechanisms can gain comprehensive insights and formulate predictions that cannot be derived from experiments or statistical data alone. Over the past decade, countless mathematical models have been developed to describe tumor growth, ranging from simple temporal population dynamic models to fully three-dimensional spatiotemporal models (Araujo and McElwain, 2006). The application of

differential equations to the specification of the proliferation and differentiation of normal SC and cancer SC has now attracted considerable attention of mathematicians and computer scientists (Adimy et al., 2008; Piotrowska et al., 2008).

However, there is currently a serious lack in the intrinsic integration of genetic, genomic and proteomic data with sophisticated mathematical models built on biological principles. Although simple statistical analysis of these data is helpful for studying the structure and function of selected signaling pathways, proteins, and drugs, it does not reveal a general dynamics and complexity of the problem and the “engine” that drives the whole biochemical machinery even in case of single cell (Piotrowska et al., 2008). An increasing recognition has been articulated about the incorporation of interdisciplinary dialogs to understand complicated cell mechanisms, network regulation, and tissue physiology. The central theme of such so-called systems biology approaches (Khalil and Hill, 2005) is to identify genes, proteins, and signaling pathways that have a functional role in the formation and expression of cancer SC by combining quantitative experimental data with mathematical modeling and computer simulation.

The objective of this article is to present a conceptual strategy for identifying genes and proteins or their expression patterns that are linked with the formation, proliferation, and programming of cancer SC. A quantitative framework will be formulated to characterize the differences in the genetic origin and biological properties between normal SC and cancerous SC, providing applied geneticists, clinical doctors, and health professionals with experimental designs and analytical tools to extract predictive information from their seemingly “chaotic” data.

## DEFINING THE GENETIC ARCHITECTURE OF CANCER

Recent advances in technology, which lead to genomic, proteomic and metabomic data collection on the human genome, now provide a powerful means of defining the genetic architecture of cancer cells. The development of any methodologies for mapping cancer genes should be based on the philosophical link of these omic data with the proliferation and differentiation of cancer SC through mathematical, statistical and engineering approaches. Below, we review several aspects that comprise the genetic architecture of cancer pathways.

### FROM GENETIC MAPPING TO GENETIC HAPLO-TYPING

For complex traits like cancer, genetic mapping is a powerful approach for illustrating their genetic architecture (Lander and Botstein, 1989; Anholt and Mackay, 2004; Li and Wu, 2010). This approach dissects complex traits into individual quantitative trait loci (QTLs) and maps them on the genome. However, a QTL detected by traditional markers may contain multiple genes that operate in a collective way, thus it is not possible to study the DNA structure, organization, and function of a QTL. A more accurate and useful approach to characterizing the underlying genetic variants is to directly analyze DNA sequences, known as quantitative trait nucleotides (QTNs), using SNPs. If a string of DNA sequence, or haplotype, for a QTN is known to increase disease risk, this risk can be intervened by inhibiting the expression of this haplotype with specialized drugs. The control of this disease can be made

more efficient if all such DNA sequences are identified throughout the entire genome. This possibility has increased dramatically with the construction of the haplotype map, or “HapMap” by SNPs (The International HapMap Consortium, 2007).

Statistical models for detecting the haplotype effects of a QTN have been derived for continuous (Liu et al., 2004; Lin and Huang, 2007) and binary traits (Cui et al., 2007). The main idea of haplotyping models is to discern the difference between unobservable diplotypes (i.e., a pair of haplotypes) from an observable unphased SNP genotype. For example, a double heterozygous genotype, AaBb, can be formed by either diplotype AB|ab or Ab|aB where | is used to separate the paternally or maternally derived haplotype. These two diplotypes, although genetically identical, can be differently responsive to cancer risk. Haplotype models allow the separation of diplotypes in the genetic control of cancer (see Liu et al., 2004).

### FROM GENETIC ACTIONS TO GENETIC INTERACTIONS

Most current approaches for cancer gene identification allow the characterization of small numbers of genes or proteins involved in cancer susceptibility. These approaches are often insufficient to study the molecular mechanisms of neoplastic processes. Cancer is a complex trait involving a network of genes that interact in a coordinated manner. These so-called genetic interactions or epistasis occur when the action of one gene measured through a molecular, cellular, or organism phenotype is modified by one or more other genes. To date, genetic interactions remain largely unknown on a large-scale in human systems. Previous reviews and assays gave excellent descriptions of the role of genetic interactions in understanding phenotypic variability (Hartman et al., 2001; Lehner et al., 2006; Boone et al., 2007). The study of genetic interactions will not only help better understand cancer susceptibility and progression, but also, and most importantly, develop novel anticancer treatments.

### FROM MENDELIAN INHERITANCE TO GENETIC IMPRINTING

Each copy of a gene, called an allele, instructs a particular task to be conducted in the cells of the body. These instructions expressed through a genetic code, i.e., a sequence of nucleotide in the DNA, guide the cells to yield a protein. In each cell of the body, there are two copies for each gene carried on a pair of maternal and paternal chromosomes, respectively. According to Mendelian inheritance, the information contained in the maternal and paternal copies of the gene is equally utilized by the cells to produce proteins. However, for certain genes, their alleles are expressed only when they transmit to a progeny through the sperm or egg. This phenomenon is called “genetic imprinting” that describes the expression of a gene relying on the paternal or maternal origin of an allele (Wilkins and Haig, 2003; Reik and Lewis, 2005; Jirtle and Skinner, 2007). Epigenetic marks that influence the expression of genes by switching the genetic information on and off may be one of the important mechanisms for genetic imprinting (Feinberg and Tycko, 2004; Sha, 2008). The imprinting modification process may be reversible in the next generation. Increasing recognition has been made that genetic imprinting plays a pivotal role in the initiation of cancer through various molecular mechanisms (Feinberg and Tycko, 2004; Secko, 2005; Sasaki and Matsui, 2008).

### FROM GENE MUTATION TO ANEUPLOIDY

Human cancer cells frequently possess large-scale chromosomal rearrangements due to chromosomal instability (CIN) or gene mutation. CIN makes whole chromosomes or large fractions of chromosomes gained or lost during cell division, resulting in an imbalance in the number of chromosomes per cell (aneuploidy), and an enhanced rate of loss of heterozygosity. Thus, the “aneuploidy hypothesis of cancer” (Stock and Bialy, 2002) proposes that the main differences between normal and abnormal (cancer) cells result from the number of genes rather than the types of genes differentially expressed, as opposed to the “gene-mutation hypothesis” (Jallepalli and Lengauer, 2001; Greenman et al., 2007; Li et al., 2010b).

Gene mutation has been a dominant hypothesis for explaining the genetics of cancer (Greenman et al., 2007), although aneuploidy was proposed to cause cancer over 100 years ago. However, growing evidence supports a role for aneuploidy in the genetic underpinning of cancer (Stock and Bialy, 2002). According to extensive work by Duesberg and his group, aneuploidy offers a simple, coherent explanation of all cancer-specific phenotypes (Duesberg, 2007; Duesberg et al., 2007). It is concluded that when aneuploidy exceeds a certain threshold it is sufficient to cause all cancer-specific phenotypes. This aneuploidy mechanism of phenotype alteration is independent of gene mutation. With a continuous confirmation of the aneuploidy hypothesis, it is crucial to quantify the genetic effects of aneuploidy loci on cancer susceptibility with the genetic data collected from the cancer genome project (Li et al., 2010a).

### FROM GENOMICS TO PROTEOMICS

With the advent of large-scale functional proteomic data (Clarke et al., 2008), additional mechanistic insights into neoplasia can be gleaned. Whole-genome association studies for cancer risk variants and somatic mutation screening projects will only provide a “part lists” of cancer genes. Transcript analyses have identified expression profiles that provide accurate prognoses for cancer patients (Fan et al., 2006). Currently, systematic mapping of protein–protein interactions leads to the birth of the so-called “interactome” mapping projects. This will elucidate the wiring diagram of protein associations in cells (Rual et al., 2005; Stelzl et al., 2005; Cohen et al., 2008). These types of genes and/or protein (gene/protein) functional relationships can be modeled together to provide better understanding and predict molecular mechanisms of neoplasia (Barabasi and Oltvai, 2004; Rhodes and Chinnaiyan, 2005; Pujana et al., 2007).

### FROM STUDYING HIGH-ORDER PHENOTYPES TO MODELING MECHANISTIC PATHWAYS

The final phenotypes of cancer size and shape are formed through a series of biological pathways. The existing strategies for genetic mapping, including functional mapping, aimed to map a biological process (Wu and Lin, 2006; Das and Wu, 2008; Li and Wu, 2010), are based directly on these high-order phenotypes, neglecting the biological, biochemical, and metabolic pathways that underlies the phenotypes. By regarding cancer formation and growth as a dynamic system, biochemical networks in the system consist of chemical reactions, such as association, dissociation, degradation,

and synthesis (Clayton et al., 2006; Hopkins, 2008). The dynamics of biochemical networks follow the rules of mechanics as well as rules governed by their own ability to organize movement and biological functions. Because a biological system is far more complex than a physical system of inert matter, we need a higher level of analysis to tackle such complexity. Mathematical models using kinetic theory may represent a way to deal with such complexity, allowing an understanding of phenomena of non-equilibrium statistical mechanics (Goldbeter, 2002). The models for system dynamics are related to the generalized Boltzmann equation, describing the population dynamics of several interacting elements (kinetic population models).

A set of non-linear ordinary or time delay differential equations have been derived to model the biological pathways and interactions between cells and between cell and surrounding environment. Biological and mechanistic behaviors of the biochemical network can be described by solving the non-linear parameters that define the differential equations. Thus, the incorporation of differential equations into genetic mapping can provide a coherent framework with which to characterize genes and the patterns of gene expression involved in mechanistic pathways.

### FROM MATHEMATIC MODELING TO STATISTICAL SOLUTION

Numerical approaches have been derived to estimate parameters for non-linear differential equations (Feng and Navaratna, 2007). Theoretical investigations of parameter estimation have been available in the applied mathematical literature (Beretta and Kuang, 2002). Increased interest has been received in the integration of statistical modeling and simulation with differential equations to study dynamic systems. In statistics, differential equations can be estimated by non-linear least squares approaches. More recently, with the availability of observational data for dynamic processes, there is a pressing need on developing more powerful statistical methods for estimating the parameter defining differential equations with noisy measurements. Ramsay et al. (2007) provided such a statistical method by combining a modified data smoothing method and the generalization of profiled estimation. The accuracy and coverage properties of parameter estimates from the new method were examined by simulation studies.

### FROM GENE–GENE INTERACTIONS TO GENOME–GENOME INTERACTIONS

Cancer arises from sporadic gene mutations, but its growth and spread will be affected by both genes from the normal and cancer genomes (Araujo and McElwain, 2006; de Araujo et al., 2008). These two different systems of genes operate interactively or epistatically to alter the course of cancer growth (de Araujo et al., 2008). Shortly after the completion of the human genome sequencing, the cancer genome is being sequenced (Kaiser, 2005). The data from these two types of genomes will provide tremendous resources for characterizing the patterns and organization of genome–genome interactions in cancer susceptibility. More recently, an interactive model has been developed to study genetic interactions of DNA sequences (or haplotypes) between host and cancer genes responsible for cancer risk using quantitative genetic principles (Li and Wu, 2009). The model is founded on a commonly used genetic association design in which a sample of related or unrelated

patients is randomly drawn from a natural human population. Each patient is typed for SNPs on normal and cancer cells and measured for cancer susceptibility. The model provides a general procedure for testing the distribution of haplotypes constructed by SNPs from host and cancer genes and the linkage disequilibria of different orders among the SNPs. The model also formulates a series of testable hypotheses about the effects of host genes, cancer genes, and their interactions on cancer susceptibility.

Li and Wu's (2009) model can be extended to detect imprinted loci based on a random set of three-generation families from a natural population by using genotyped SNPs. This design provides a pathway for characterizing the effects of imprinted genes on a complex disease at different generations and testing transgenerational changes of imprinted effects (Youngson and Whitelaw, 2008; Li et al., 2011b). The design is integrated with population and cytogenetic principles of gene segregation and transmission from the previous generation to the next.

Currently used genome-wide association studies (GWAS) allow the scan of functional or causal polymorphisms from 300,000 to 1 million SNPs (Pearson and Manolio, 2008). Genetic mapping has developed to a point at which a comprehensive analysis of all the markers that cover the genetic map of the genome can be performed to search for the chromosomal distribution of all possible QTLs or QTNs. Some basic work in GWAS of QTLs has been initiated in recent years, although it is still full of challenges. Several statistical models for analyzing all SNPs through lasso and other variable selection approaches have been developed (Li et al., 2011a). These models shrink the effects of most SNPs toward zero to provide sparse estimates of whole-genome associations.

### MODELING DYNAMIC GENETIC CONTROL

Mathematical models have been increasingly used to describe the population dynamics of cancer SC and their differentiation into cancer cells (Solyani et al., 1995; Ganguly and Puri, 2006; Enderling et al., 2007; Adimy et al., 2008; Michor, 2008; Piotrowska et al., 2008). There is strong evidence that genes due to mutations in normal SC are involved in the regulation of the self-renewal of cancer SC for cancers (Al-Hajj and Clarke, 2004; Clarke, 2004). Statistical models are needed to map those genes from mathematical aspects of cancer SC.

### ORDINARY DIFFERENTIAL EQUATIONS (ODE)

A large system of non-linear ordinary differential equations (ODE) has been derived to the self-renewing capacity of cancer SC (Ganguly and Puri, 2006; Piotrowska et al., 2008). These equations were founded on the principle according to which cancers result from mutations in normal SC, early progenitor (EP) cells and even mature cells (MC). The model for brain cancer cells by Ganguly and Puri (2006) defines seven main types of cells – SC, EP cells, late progenitor (LP) cells, MC, abnormal stem cells (SCA), abnormal early progenitor cells (EP<sub>A</sub>), and abnormal progeny (AP; tumor) cells. Each cell type is considered as a separate model compartment, whose cell population growth is modeled by considering individual rate expressions (Figure 1). SCs can self-renew with a probability  $P_{SC}$ , in which both daughter cells retain stem cell features, or differentiate and transfer to the EP compartment. Stem cell DNA may mutate during the replication with probability  $M_{SC}$  by which the daughter cell inheriting the mutated gene is

transferred into the SCA population. EP cells, as well as EP<sub>A</sub> cells, experience only a limited number ( $K$ ) of self-renewal steps. Thus, cells with identical self-renewal capacity are grouped into  $K$  sub-compartments, although cells belonging to the  $k$ th ( $k = 1, \dots, K$ ) compartment cannot self-renew any more. If EP<sub>k</sub> cells undergo cell division, they self-renew into a subgroup EP<sub>k+1</sub> with a given probability  $P_{EP}$ , which is assumed to be equal for all subpopulations. The same is true for EP<sub>Ak</sub> which self-renews into EP<sub>Ak+1</sub> with a probability  $P_{EP,A}$ . Dividing cells that do not supply the EP<sub>i+1</sub> or EP<sub>Ai+1</sub> compartment differentiate into LP or AP (in case of abnormal cells) cells, respectively. Furthermore, at each division EP<sub>i</sub> cells are subject to mutations as defined by mutation probability  $M_{EP}$ , which is assumed to be identical for each EP<sub>i</sub> compartment. Naturally, the AP compartment is supplied by EP<sub>A</sub> cells. Finally, cells that reach the MC or AP compartment die due to apoptosis (Piotrowska et al., 2008).

Ganguly and Puri (2006) described the rates of joint changes of different types of cell populations by a system of differential equations expressed as

Stem cell compartment

$$\dot{N}_{SC} = \omega_{SC} (2P_{SC} - 1) N_{SC} - \omega_{SC} M_{SC} N_{SC}$$

Early progenitor cell compartment

$$\begin{cases} \dot{N}_{EP1} = 2\omega_{SC} (1 - P_{SC}) N_{SC} - \omega_{EP} N_{EP1} \\ \dot{N}_{EPk} = 2\omega_{EP} (2P_{EP} - M_{EP}) N_{EP(k-1)} - \omega_{EP} N_{EPk}, \\ \text{for } k = 2, \dots, K \end{cases}$$

Late progenitor cell compartment

$$\begin{aligned} \dot{N}_{LP} = Z^{\text{in}} \left[ \sum_{k=1}^{K-1} 2\omega_{EP} (1 - P_{EP}) N_{EPk} + 2\omega_{EP} N_{EPK} \right] \\ - \omega_{LP} N_{LP} \end{aligned}$$

Mature cell compartment

$$\dot{N}_{MC} = Z_{LP}^{\text{out}} - \omega_{0,MC} N_{MC}$$

Abnormal stem cell compartment

$$\dot{N}_{SC,A} = \omega_{SC} (2P_{SC} - 1) N_{SC,A} + \omega_{SC} M_{SC} N_{SC}$$

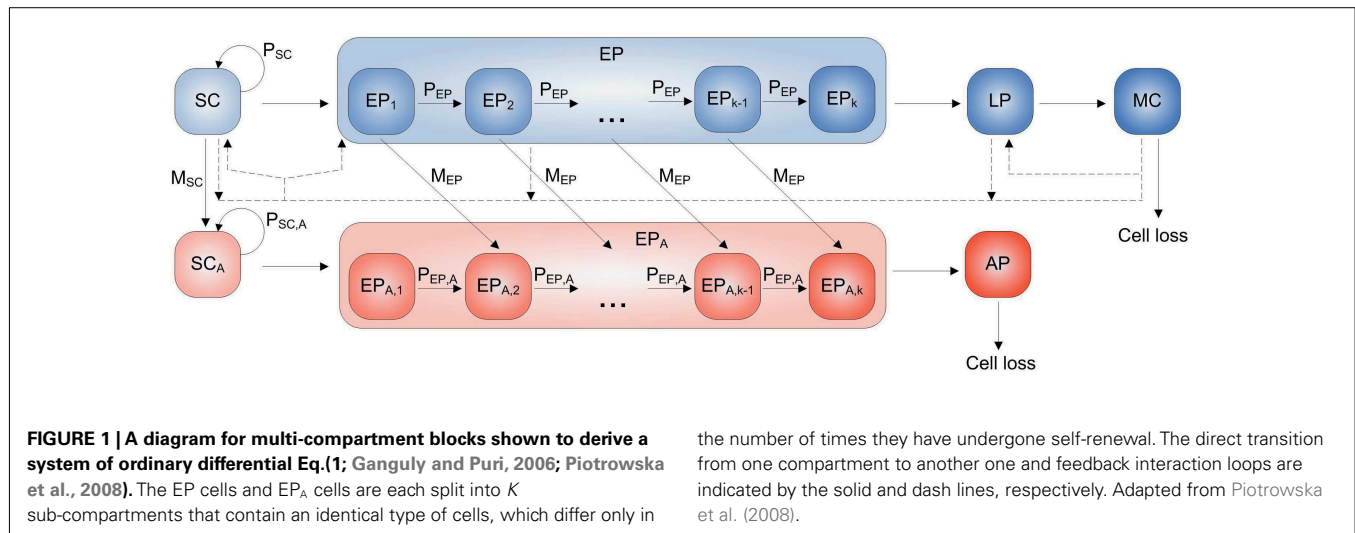
Abn. early progenitor cell compart.

$$\begin{cases} \dot{N}_{EPA1} = 2\omega_{SC} (1 - P_{SC}) N_{SC,A} - \omega_{EP} N_{EPA1} \\ \dot{N}_{EPAk} = 2\omega_{EP} P_{EP} N_{EPA,k-1} - \omega_{EP} N_{EPAk} + \omega_{EP} N_{EP(k-1)} M_{EP}, \\ \text{for } k = 2, \dots, K \end{cases}$$

Abnormal progeny compartment

$$\dot{N}_{AP} = Z_{EPA}^{\text{out}} - \omega_{0,AP} N_{AP} \tag{1}$$

where  $N$  is the number of cells;  $\omega_{SC}$ ,  $\omega_{EP}$ , and  $\omega_{LP}$  are the cell division rates with SC, EP, and LP compartments, respectively;  $P_{SC} = P_{SC}(N_{SC}, N_{EP}, N_{LP}, N_{MC}) \in [P_{\text{min},SC}, P_{\text{max},SC}]$  is the self-renewal probability;  $Z^{\text{in}}$  is the factor by which the efflux from the EP compartment is amplified as soon as the cells enter the LP compartment;  $Z_{LP}^{\text{out}} = Z^{\text{out}} \omega_{LP} N_{LP}$  is the efflux from the LP compartment, i.e., the generation rate of MC, where  $Z^{\text{out}}$  is the factor by which the cells are further amplified immediately before leaving the LP compartment;  $\omega_{0,MC}$  is the cell death or apoptotic rate;  $Z_{EPA}^{\text{out}} = \sum_{k=1}^{K-1} 2\omega_{EP} (1 - P_{EP}) N_{EPAk} + 2\omega_{EP} N_{EPK}$  is the outflux of abnormal EP cells; and  $\omega_{0,AP}$  is the AP cell death or



apoptotic rate. Cell populations in the EP and EP<sub>A</sub> compartments are determined by the output equations  $N_{EP} = \sum_{k=1}^K N_{EP_k}$  and  $N_{EP_A} = \sum_{k=1}^K N_{EP_{A,k}}$ , respectively.

Cell division rates  $\omega_{SC} = (\alpha_{SC}/\tau_{SC})\ln 2$ ,  $\omega_{EP} = (\alpha_{EP}/\tau_{EP})\ln 2$ , and  $\omega_{LP} = (\alpha_{LP}/\tau_{LP})\ln 2$  depend on the mitotic fraction ( $\alpha$ ) and the cell cycle time ( $\tau$ ), i.e., the time required to complete a full cycle of cell division. The total time  $\tau_{LP}$  spent by a cell in the LP compartment is the sum of the cell generation time ( $\tau_g = \tau_{LP}/\alpha_{LP}\ln 2$ ) and the cell maturation time  $\tau_m$ . The two factors  $Z^{\text{in}}$  and  $Z^{\text{out}}$  are selected such that  $Z^{\text{in}} \times Z^{\text{out}} = 2^n$ , where  $Z^{\text{in}} = (2^n - 1)\tau_g/\tau_{LP} + 2^n\tau_m/\tau_{LP}$  and  $n$  is the number of stages of cell division of the LP compartment before transforming into MC. All in all, the dynamic system of cancer formation from cancer SC can be quantified by the group of ODEs (1) with a set of parameters, cell cycle times ( $\tau_{SC}$ ,  $\tau_{EP}$ ,  $\tau_{LP}$ ), cell maturation time for LP compartment  $\tau_m$ , self-renewal probabilities ( $P_{SC}$ ,  $P_{EP}$ ), mitotic fractions ( $\alpha_{SC}$ ,  $\alpha_{EP}$ ,  $\alpha_{LP}$ ), number of EP cell self-renewals  $K$ , the number of mitotic cycles  $n$ , and death rates ( $\omega_{0,MC}$ ,  $\omega_{0,AP}$ ), and the probabilities of occurrence of oncogenic mutations during DNA transcription ( $M_{SC}$ ,  $M_{EP}$ ).

The solution of these differential equations is not known in closed, analytic form, and hence must be computed approximately by means of algorithms and software from numerical mathematics and scientific computing. Savcenco (2009) proposed a multirate time stepping technique for a system of ODEs. This method uses large time steps for slowly varying components, and small steps for rapidly varying ones. Numerical experiments confirm that the efficiency of time integration methods can be significantly improved by using multirate methods.

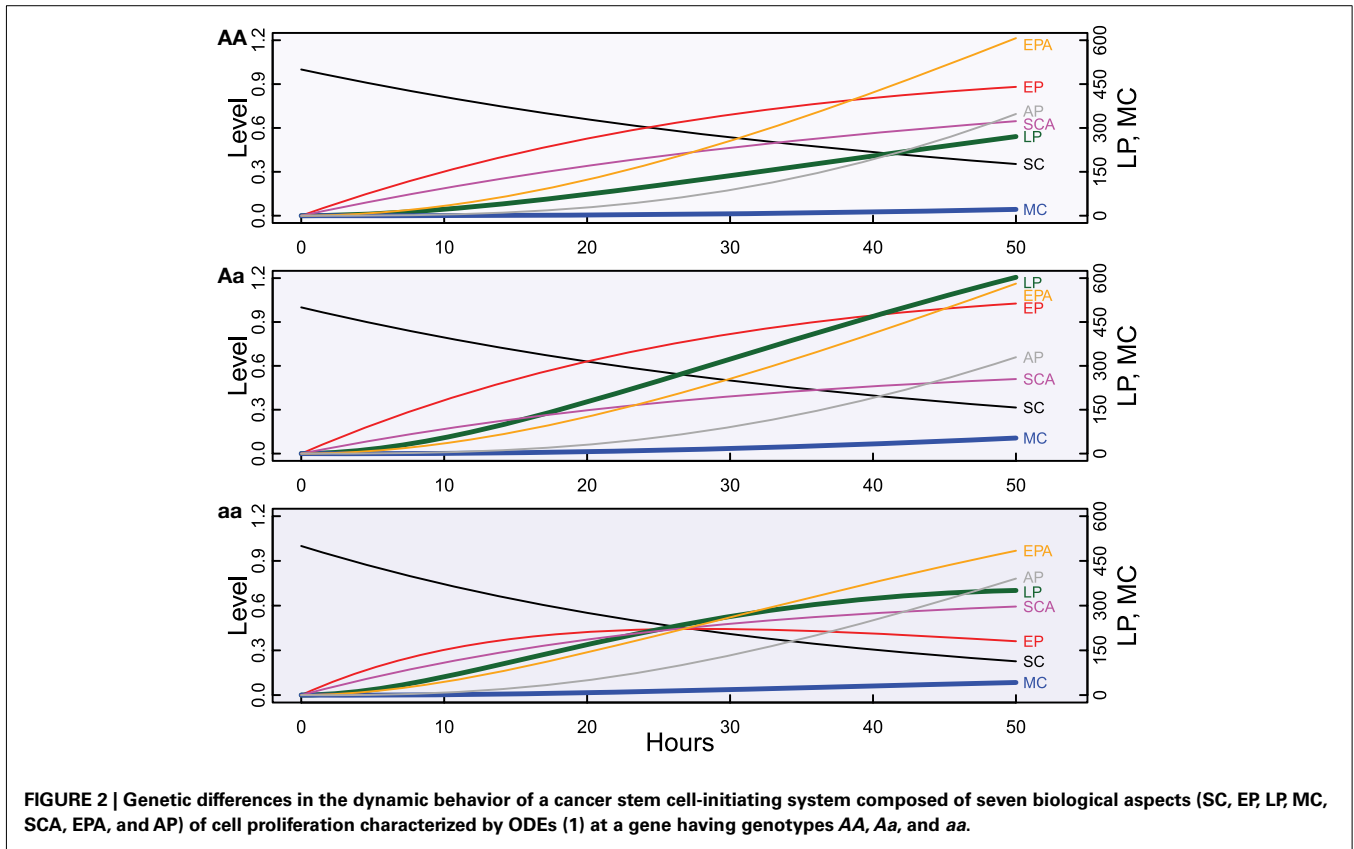
Given the fact that there is a substantial variability in cancer formation from cancer SC, we integrate the differential Eq. (1) into a genetic mapping framework (Fu et al., 2010; Luo et al., 2010). By estimating ODE parameters for individual genotypes at a (or a set of) SNP, we can test whether there exist significant loci associated with the capacity of cancer SC to develop into malignant cancers and how these significant loci affect various proliferation and apoptotic processes within different cell compartments. Consider a QTL with three genotypes AA, Aa, and aa bearing ODE

parameters ( $\tau_{SC}$ ,  $\tau_{EP}$ ,  $\tau_{LP}$ ;  $\tau_m$ ;  $P_{SC}$ ,  $P_{EP}$ ;  $\alpha_{SC}$ ,  $\alpha_{EP}$ ,  $\alpha_{LP}$ ;  $K$ ;  $n$ ;  $\omega_{0,MC}$ ,  $\omega_{0,AP}$ ;  $M_{SC}$ ,  $M_{EP}$ ) = (8 h, 8 h, 8 h; 40 h; 0.5, 0.5; 0.4, 0.4; 0.35, 5; 6; 0.01, 0.01; 0.6, 0.6), (9 h, 9 h, 9 h; 45 h; 0.45, 0.45; 0.5, 0.45, 0.45; 6; 7; 0.02, 0.02; 0.5, 0.5), and (7 h, 8 h, 9 h; 40 h; 0.45, 0.3; 0.4, 0.45, 0.5; 7; 7; 0.03, 0.03; 0.65, 0.65), respectively. **Figure 2** provides the dynamic behavior of each cell type separately for different QTL genotypes. The integrated mapping models allow the formulation of numerous biologically meaningful hypothesis tests about the genetic control of each pathway in the system.

#### DELAY DIFFERENTIAL EQUATIONS (DDE)

All blood cells arise from a common origin in the bone marrow, the hematopoietic SC. These SC can differentiate into one of three major cell lines: the leukocytes, the platelets, and the erythrocytes. Because of genetic mutations, the hematopoietic SC become leukemia SC, endowed with the capacity to self-renew and to generate many leukaemic blasts (Viale et al., 2009). Mackey and colleagues (Colijn and Mackey, 2005a,b; Piotrowska et al., 2008) published a series of quantitative models that specify four distinct compartments representing the hematopoietic SC and the circulating leukocytes, platelets, and erythrocytes. Typically all of the three differentiated cell types have the same oscillation period, but the relation of the oscillation mean and amplitude to the normal levels is variable. Due to chronic features of these cells, a system of delay differential equations (DDEs) has been derived to specify the dynamics of the system. A number of mathematicians studied the stability properties of these delay non-linear differential equations and identified the supercritical Hopf bifurcation and a saddle-node bifurcation of limit circles (Bernard et al., 2003; Kirk et al., 2008). It is concluded that the major cause of the oscillation in cyclic neutropenia is an increase rate of apoptosis of neutrophil precursors leading to a destabilization of hematopoietic stem cell compartments (Crauste, 2006; Feng and Navaratna, 2007; Verdugo and Rand, 2008).

We studied the influences of different DDE parameters on the period and amplitude of oscillations if a specific set of DDEs is given (Fu et al., 2011), which helps to obtain a solution of the DDEs. In Fu et al. (2011), we have developed a framework for



**FIGURE 2 | Genetic differences in the dynamic behavior of a cancer stem cell-initiating system composed of seven biological aspects (SC, EP, LP, MC, SCA, EPA, and AP) of cell proliferation characterized by ODEs (1) at a gene having genotypes AA, Aa, and aa.**

incorporating DDEs into the functional mapping of a periodic system (Liu et al., 2007). This framework can be readily extended to study the genetic mapping of chronic myelogenous leukemia by incorporating Colijn and Mackey’s (2005a,b) DDEs. As shown in Wu et al. (2004), the genetic control mechanisms for the biochemical connections between differentiated cells and SC will be formulated in a quantitative and testable ways.

**SYSTEMS MAPPING**

Systems mapping is a computational model that views a complex phenotype as a dynamic system, dissects it into its underlying components, coordinates different components in terms of biological laws through mathematical equations, and maps specific genes that mediate each component and its connection with other components (Wu et al., 2011). As a bottom–top model, systems approach can identify specific QTLs that govern the developmental interactions of various components giving rise to the function and behavior of the system. By estimating and testing mathematical parameters that specify the system, systems mapping enables the prediction or alteration of the physiological status of a phenotype based on the underlying genetic control mechanisms.

**Mixture model**

Systems mapping embeds a system of differential equations, like (1), into a genetic mapping setting constructed by a segregating population. Genetic mapping uses a mixture model-based likelihood to estimate genotype-specific parameters by assuming  $J$  QTL genotypes. For the mapping population of  $n$  members with

marker information ( $\mathbf{M}$ ) and phenotypic data for different types of cancer SC ( $\mathbf{Y}$ ), we formulate the likelihood as

$$L(\mathbf{M}, \mathbf{Y}) = \prod_{i=1}^n \sum_{j=1}^J \omega_{ji} f_j(\mathbf{Y}_i; \Theta_j, \Psi) \tag{2}$$

where  $\mathbf{Y}_i = (\mathbf{Y}_{1i}, \dots, \mathbf{Y}_{7i})$  is the phenotypic vector for individual  $i$  containing seven types of cancer SC, SC ( $\mathbf{Y}_1$ ), EP ( $\mathbf{Y}_2$ ), LP ( $\mathbf{Y}_3$ ), MC ( $\mathbf{Y}_4$ ), SCA ( $\mathbf{Y}_5$ ), EPA ( $\mathbf{Y}_6$ ), AP ( $\mathbf{Y}_7$ ) [see (1)], whose  $k$ th type is a time-dependent vector measured at  $T$  time points, i.e.,  $\mathbf{Y}_k = (Y_k(1), \dots, Y_k(T))$  ( $k = 1, \dots, 7$ );  $\omega_{ji}$  is the mixture proportion representing the conditional probability of QTL genotype  $j$  given the marker genotype of individual  $i$ ; and  $f_j(\mathbf{Y}_i; \Theta_j, \Psi)$  is a multivariate normal distribution with expected mean vector for individual  $i$  that belongs to QTL genotype  $j$ ,

$$\boldsymbol{\mu}_{j/i} = (\boldsymbol{\mu}_{1j/i}(1); \dots; \boldsymbol{\mu}_{7j/i}(T)) \equiv (\boldsymbol{\mu}_{1j/i}(1), \dots, \boldsymbol{\mu}_{7j/i}(1); \dots; \boldsymbol{\mu}_{1j/i}(T), \dots, \boldsymbol{\mu}_{7j/i}(T)) \tag{3}$$

and covariance matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_1 & \dots & \boldsymbol{\Sigma}_{17} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{71} & \dots & \boldsymbol{\Sigma}_7 \end{pmatrix} \tag{4}$$

with  $\boldsymbol{\Sigma}_k$  being a  $(T \times T)$  covariance matrix for cell type  $k$  and  $\boldsymbol{\Sigma}_{kk'} = \boldsymbol{\Sigma}_{k'k}^T$  being a  $(T \times T)$  covariance matrix between cell types  $k$  and  $k'$ .

**Table 1 | The estimation of parameters by a dynamic model for mapping cancer stem cell proliferation from a simulated mapping population of 500 individuals by assuming heritability  $H^2=0.1$**

	AA		Aa		aa	
	True	Est (SE)	True	Est (SE)	True	Est (SE)
$\tau_{SC}$	9	9.354 (2.175)	7	8.014 (1.983)	8	9.855 (5.871)
$\tau_{EP}$	9	8.181 (2.351)	8	7.716 (1.397)	8	7.961 (3.180)
$\tau_{LP}$	9	9.568 (2.142)	9	9.533 (3.099)	8	8.664 (14.374)
$\tau_m$	45	45.169 (1.219)	40	39.834 (1.116)	40	39.482 (3.349)
$P_{SC}$	0.45	0.450 (0.007)	0.45	0.441 (0.020)	0.5	0.475 (0.081)
$P_{EP}$	0.45	0.451 (0.030)	0.3	0.292 (0.189)	0.5	0.465 (0.618)
$\alpha_{SC}$	0.5	0.517 (0.142)	0.4	0.433 (0.133)	0.4	0.471 (0.264)
$\alpha_{EP}$	0.45	0.422 (0.113)	0.45	0.470 (0.158)	0.4	0.560 (0.881)
$\alpha_{LP}$	0.45	0.469 (0.069)	0.5	0.506 (0.082)	0.35	0.349 (1.070)
$\omega_{0,SC}$	0.02	0.023 (0.017)	0.03	0.030 (0.017)	0.01	0.219 (1.135)
$\omega_{0,AP}$	0.02	0.025 (0.013)	0.03	0.033 (0.011)	0.01	0.061 (0.227)
$M_{SC}$	0.5	0.507 (0.061)	0.65	0.728 (0.237)	0.6	1.028 (2.157)
$M_{EP}$	0.5	0.494 (0.037)	0.65	0.635 (0.112)	0.6	0.560 (0.805)
$K$	6.000	6.000 ( $4.0 \times 10^{-9}$ )	7.000	7.000 ( $5.5 \times 10^{-09}$ )	5.000	5.000 ( $2.7 \times 10^{-9}$ )
$n$	7.000	7.000 ( $1.2 \times 10^{-8}$ )	7.000	7.000 ( $9.2 \times 10^{-10}$ )	6.000	6.000 ( $1.2 \times 10^{-8}$ )
<b>AR (1)</b>						
$\sigma_1$	0.182	0.181 (0.005)				
$\sigma_2$	0.812	0.810 (0.018)				
$\sigma_3$	2.781	2.767 (0.056)				
$\sigma_4$	0.350	0.350 (0.005)				
$\sigma_5$	0.122	0.122 (0.003)				
$\sigma_6$	0.354	0.355 (0.007)				
$\sigma_7$	0.138	0.130 (0.045)				
$\rho_1$	0.8	0.797 (0.011)				
$\rho_2$	0.75	0.745 (0.014)				
$\rho_3$	0.65	0.646 (0.016)				
$\rho_4$	0.55	0.551 (0.013)				
$\rho_5$	0.75	0.747 (0.011)				
$\rho_6$	0.65	0.650 (0.015)				
$\rho_7$	0.55	0.543 (0.017)				
<b>ALLELE FREQUENCIES AND LINKAGE DISEQUILIBRIUM (LD)</b>						
$\rho(\text{Marker})$	0.6	0.598 (0.017)				
$Q$ (QTL)	0.7	0.679 (0.054)				
$D$ (LD)	0.05	0.053 (0.016)				

The means and standard errors of parameter estimates are calculated from 100 simulation replicates.

**Table 2 | The estimation of parameters by a dynamic model for mapping cancer stem cell proliferation from a simulated mapping population of 500 individuals by assuming heritability  $H^2=0.4$**

	AA		Aa		aa	
	True	Est (SE)	True	Est (SE)	True	Est (SE)
$\tau_{SC}$	9	9.672 (2.230)	7	7.827 (1.831)	8	8.542 (2.329)
$\tau_{EP}$	9	9.069 (2.688)	8	8.334 (2.082)	8	7.681 (2.143)
$\tau_{LP}$	9	9.032 (0.741)	9	9.137 (0.799)	8	8.647 (3.708)
$\tau_m$	45	45.304 (1.594)	40	40.340 (1.660)	40	40.110 (1.536)
$P_{SC}$	0.45	0.450 (0.003)	0.45	0.450 (0.004)	0.5	0.500 (0.004)
$P_{EP}$	0.45	0.451 (0.011)	0.3	0.296 (0.033)	0.5	0.502 (0.033)

(Continued)

Table 2 | Continued

	AA		Aa		aa	
	True	Est (SE)	True	Est (SE)	True	Est (SE)
$\alpha_{SC}$	0.5	0.534 (0.122)	0.4	0.450 (0.115)	0.4	0.417 (0.101)
$\alpha_{EP}$	0.45	0.455 (0.132)	0.45	0.468 (0.115)	0.4	0.399 (0.142)
$\alpha_{LP}$	0.45	0.453 (0.028)	0.5	0.506 (0.031)	0.35	0.368 (0.103)
$\omega_{0,SC}$	0.02	0.020 (0.006)	0.03	0.030 (0.006)	0.01	0.013 (0.035)
$\omega_{0,AP}$	0.02	0.020 (0.004)	0.03	0.030 (0.003)	0.01	0.013 (0.013)
$M_{SC}$	0.5	0.504 (0.022)	0.65	0.648 (0.043)	0.6	0.614 (0.079)
$M_{EP}$	0.5	0.501 (0.013)	0.65	0.650 (0.034)	0.6	0.601 (0.036)
$K$	6	6.000 ( $1.1 \times 10^{-09}$ )	7	7.000 ( $3.8 \times 10^{-10}$ )	5	5.000 ( $7.9 \times 10^{-10}$ )
$n$	7	7.000 ( $5.8 \times 10^{-10}$ )	7	7.000 ( $1.3 \times 10^{-09}$ )	6	6.000 ( $9.8 \times 10^{-10}$ )
<b>AR (1)</b>						
$\sigma_1$	0.074	0.013 (0.073)				
$\sigma_2$	0.331	0.330 (0.008)				
$\sigma_3$	1.135	1.135 (0.021)				
$\sigma_4$	0.143	0.143 (0.003)				
$\sigma_5$	0.049	0.049 (0.001)				
$\sigma_6$	0.145	0.145 (0.003)				
$\sigma_7$	0.057	0.051 (0.025)				
$\rho_1$	0.8	0.801 (0.011)				
$\rho_2$	0.75	0.747 (0.014)				
$\rho_3$	0.65	0.650 (0.014)				
$\rho_4$	0.55	0.547 (0.019)				
$\rho_5$	0.75	0.750 (0.012)				
$\rho_6$	0.65	0.650 (0.015)				
$\rho_7$	0.55	0.549 (0.019)				
<b>ALLELE FREQUENCIES AND LINKAGE DISEQUILIBRIUM (LD)</b>						
$p(\text{Marker})$	0.6	0.601 (0.015)				
$Q$ (QTL)	0.7	0.701 (0.014)				
$D$ (LD)	0.05	0.052 (0.010)				

The means and standard errors of parameter estimates are calculated from 100 simulation replicates.

Traditional approaches estimate every element in (3) and (4), but systems mapping specifies and models vector  $\mu_{ji}$  by ODE parameters  $\Theta_j$  and covariance matrix  $\Sigma$  by  $\Psi$ . Next, we show how to model and estimate  $\omega_{ji}$ ,  $\Theta_j$ , and  $\Psi$ .

### Architectural and dynamic modeling

Mixture proportion  $\omega_{ji}$  relates to the pattern of QTL segregation whose optimal description relies on the genetic architecture of phenotypic traits. In the preceding section, we review several key aspects of the genetic architecture of cancer SC, which can be embedded into systems mapping (2).  $\omega_{ji}$  can be expressed at the individual locus level or haplotype level (Liu et al., 2004), and can be used to model genetic main effects or epistatic interactions that occur within and between genomes (Li and Wu, 2009).  $\omega_{ji}$  can also consider genetic imprinting and epigenetic effects by incorporating the parental information of mapping individuals. In particular, genetic causes of cancer, like gene mutation or aneuploidy, can be tested through  $\omega_{ji}$ , providing a way to test the hypothesis of cancer formation.

The dynamic behavior of cancer phenotypes can be described by various differential equations. We have incorporated commonly used mathematical tools, like the Runge–Kutta fourth order

algorithm, to provide the estimates of mathematical parameters for individual QTL genotypes contained within a mixture-model framework (Fu et al., 2010). Mathematical solution for delay differential equations has been discussed and used to map clock genes for a biological system (Fu et al., 2011). For longitudinal data, we can use structural approaches to model the covariance matrix for longitudinal traits (Zimmerman and Nunez-Anton, 2001; Zhao et al., 2005; Yap et al., 2009). These approaches include (1) parametric stationary, (2) parametric non-stationary, (3) non-parametric, and (4) semiparametric models. Each of these approaches has advantages and disadvantages in computing efficiency, flexibility, and power.

### Numerical simulation

We performed a simple simulation studies to test the statistical behavior of systems mapping for cancer SC. Fu et al. (2010) provided much more detailed information about the test and validation of systems mapping. The simulation assumes a mapping population of different heritabilities at a middle stage of cancer stem cell growth. Consider a SNP marker which is associated with a putative QTL (with three genotypes AA, Aa, and aa) through linkage disequilibrium (LD). Given allele frequencies of the marker



and QTL and their LD, joint marker and phenotypic data were simulated. The genotypic mean vectors were modeled by ODE (1) using parameters  $\Theta_j = (\tau_{SCj}, \tau_{EPj}, \tau_{LPj}; P_{SCj}, P_{EPj}; \alpha_{SCj}, \alpha_{EPj}, \alpha_{LPj}; K_j; n_j; \omega_{0,MCj}, \omega_{0,APj}; M_{SCj}, M_{EPj})$ , whereas the covariance matrix was structured by the first-order autoregressive [AR(1)] model with correlation and variance parameters for cell type  $k$ ,  $\Psi_k = (\rho, \sigma^2)$ . We assume no residual correlation between different cell types.

**Tables 1 and 2** provide the results about the estimates of ODE parameters for individual QTL genotypes and covariance-structural parameters, along with marker and QTL segregation parameters. The parameters for marker and QTL allele frequencies and LD can well be estimated, although missing QTL information is inferred using the mixture model. The AR(1) parameters can also well be estimated, partly because of their simple structure for covariance modeling. The precision of ODE parameter estimation heavily rely on heritability and sample size. For a modest heritability (say 0.1), some ODE parameters, like  $\tau_{LP}$ , in a small group of QTL genotype ( $aa$ ) were not precisely estimated (**Table 1**). Given its allele frequency 0.3, the number of genotype  $aa$  is less than 50. However, when heritability increases to 0.4, all ODE parameters including ones in small-sized genotype  $aa$  can be reasonably estimated (**Table 2**). The results from the simulation study suggest that the increase of trait heritability through reducing phenotyping errors is an important measure for improving the precision of parameter estimation.

## DISCUSSION

Tremendous investments have been made on cancer genome projects in order to make a complete landscape of mutations in tumors accessible. Although the importance of cataloging genome variations is well recognized, there are obvious difficulties in bridging the gaps between high-throughput sequencing information and molecular mechanisms of cancer evolution. In this article,

we argue that the integration of functional and systems mapping (Wu and Lin, 2006; Li and Wu, 2010; Wu et al., 2011) with mathematical models for describing the origin and proliferation of cancer SC (Colijn and Mackey, 2005a,b; Ganguly and Puri, 2006; Enderling et al., 2007; Adimy et al., 2008; Michor, 2008; Piotrowska et al., 2008) may shed light on a better understanding of cancer genetics. We particularly emphasize the importance of integrating the latest discoveries of cancer genetics in a diversity of areas from haplotype effects to gene mutation or aneuploidy to genetic imprinting into statistical functional mapping models.

We have constructed a general framework for modeling the genetic control of the dynamics of cancer stem cell populations based on the general characterization of a tumor. The model uses specific differential equation parameters to describe different stages of tumor progression and test the magnitude and patterns of genetic effects on the proliferation of cancer SC. The model should be extended to model differences in proliferation dynamics of normal SC and cancer SC, which facilitates the diagnosis of cancer initiation and malignant progression from individuals' genetic information.

The model allows the target of cancer SC based on genetic data. By coupling with "omics" data, the model may offer additional therapeutic window for success in disease elimination. Also, mathematical modeling of gene regulatory networks in cancer SC may provide insight into the underlying mechanism for disease prevention and for the provision of novel insight into new therapeutic formulations. A more sophisticated model that takes into account multi-step processes with several feedback functions and therapeutic responses of cancer interventions would be very beneficial in future studies.

## ACKNOWLEDGMENTS

This work is partly supported by NIH/UL1RR0330184.

## REFERENCES

- Adimy, M., Angulo, O., Crauste, F., and Lopez-Marcos, J. C. (2008). Numerical integration of a mathematical model of hematopoietic stem cell dynamics. *Comput. Math. Appl.* 56, 594–600.
- Al-Hajj, M., and Clarke, M. F. (2004). Self-renewal and solid tumor cells. *Oncogene* 23, 7274–7282.
- Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., and Clarke, M. F. (2003). Prospective identification of tumorigenic breast cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3983–3988.
- Alison, M. R. (2005). Liver stem cells: implications for hepatocarcinogenesis. *Stem Cell Rev.* 1, 253–260.
- Anderson, A. R. A., and Quaranta, V. (2008). Integrative mathematical oncology. *Nat. Rev. Cancer* 8, 227–233.
- Anholt, R. R., and Mackay, T. F. C. (2004). Quantitative genetic analyses of complex behaviours in *Drosophila*. *Nat. Rev. Genet.* 5, 838–849.
- Araujo, R. P., and McElwain, D. L. S. (2006). The role of mechanical host-tumour interactions in the collapse of tumour blood vessels and tumour growth dynamics. *J. Theor. Biol.* 238, 817–827.
- Ayers, M., Symmans, W. F., Stec, J., Damokosh, A. I., Clark, E., Hess, K., Lecoche, M., Metivier, J., Booser, D., Ibrahim, N., Valero, V., Royce, M., Arun, B., Whitman, G., Ross, J., Sneige, N., Hortobagyi, G. N., and Pusztai, L. (2004). Gene expression profiles predict complete pathologic response to neoadjuvant paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide chemotherapy in breast cancer. *J. Clin. Oncol.* 22, 2284–2293.
- Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113.
- Baumann, M., Krause, M., and Hill, R. (2008). Exploring the role of cancer stem cells in radioresistance. *Nat. Rev. Cancer* 8, 545–554.
- Beretta, E., and Kuang, Y. (2002). Geometric stability switch criteria in delay differential systems with delay dependent parameters. *SIAM J. Math. Anal.* 33, 1144–1165.
- Bernard, S., Belair, J., and Mackey, M. (2003). Oscillations in cyclical neutropenia: New evidence based on mathematical modeling. *J. Theor. Biol.* 223, 283–298.
- Bhanot, G., Alexe, G., Levine, A. J., and Stolovitzky, G. (2005). Robust diagnosis of non-Hodgkin lymphoma phenotypes validated on gene expression data from different laboratories. *Genome Inform.* 16, 233–244.
- Boone, C., Bussey, H., and Andrews, B. J. (2007). Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* 8, 437–449.
- Brown, M. D., Gilmore, P. E., Hart, C. A., Samuel, J. D., Ramani, V. A., George, N. J., and Clarke, N. W. (2007). Characterization of benign and malignant prostate epithelial Hoechst 33342 side populations. *Prostate* 67, 1384–1396.
- Clarke, M. F. (2004). At the root of brain cancer. *Nature* 432, 281–282.
- Clarke, M. F., and Fuller, M. (2006). Stem cells and cancer: two faces of eve. *Cell* 124, 1111–1115.
- Clarke, R., Renshaw, H. W., Wang, A., Xuan, J., Liu, M. C., Gehan, E. A., and Wang, Y. (2008). The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. *Nat. Rev. Cancer* 8, 37–49.
- Clayton, T. A., Lindon, J. C., Cloarec, O., Antti, H., Charuel, C., Hanton, G., Provost, J.-P., Le Net, J.-L., Baker, D., Walley, R. J., Everett, J. R., and Nicholson, J. K. (2006). Pharmacometabonomic phenotyping and personalized drug treatment. *Nature* 440, 1073–1077.
- Cohen, A. A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z.,

- Cohen, L., Danon, T., Perzov, N., and Alon, U. (2008). Dynamic protoeconomics of individual cancer cells in response to a drug. *Science* 322, 1511–1516.
- Colijn, C., and Mackey, M. C. (2005a). A mathematical model of hematopoiesis: periodic chronic myelogenous leukemia, part I. *J. Theor. Biol.* 237, 117–132.
- Colijn, C., and Mackey, M. C. (2005b). A mathematical model of hematopoiesis: cyclical neutropenia, part II. *J. Theor. Biol.* 237, 133–146.
- Crauste, F. (2006). Global asymptotic stability and Hopf bifurcation for a blood cell production model. *Math. Biosci. Eng.* 3, 325–346.
- Cui, Y. H., Sun, K. L., Fan, X., and Wu, R. L. (2007). Mapping nucleotide sequences that encode complex binary disease traits with HapMap. *Curr. Genomics* 8, 307–322.
- Das, K., and Wu, R. L. (2008). A statistical model for the identification of genes governing the incidence of cancer with age. *Theor. Biol. Med. Model.* 5, 7.
- de Araujo, S. P. S., Maciag, P. C., Ribeiro, K. B., Petzl-Erler, M. L., Franco, E. L., and Villa, L. L. (2008). Interaction between polymorphisms of the human leukocyte antigen and HPV-16 variants on the risk of invasive cervical cancer. *BMC Cancer* 8, 246. doi:10.1186/1471-2407-8-246
- Dick, J. E. (2003). Breast cancer stem cells revealed. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3547–3549.
- Dontu, G., Al-Hajj, M., Abdallah, W. M., Clarke, M. F., and Wicha, M. S. (2003). Stem cells in normal breast development and breast cancer. *Cell Prolif.* 36, 59–72.
- Duesberg, P. (2007). Chromosomal chaos and cancer. *Sci. Am.* 296, 52–59.
- Duesberg, P., Li, R., Fabarius, A., and Hehlmann, R. (2007). Orders of magnitude change in phenotype rate caused by mutation. *Cell. Oncol.* 29, 71–72.
- Enderling, H., Chaplain, M. A., Anderson, A. R., and Vaidya, J. S. (2007). A mathematical model of breast cancer development, local treatment and recurrence. *J. Theor. Biol.* 246, 245–259.
- Fan, C., Oh, D. S., Wessels, L., Weigelt, B., Nuyten, D. S., Nobel, A. B., van't Veer, L. J., and Perou, C. M. (2006). Concordance among gene-expression-based predictors for breast cancer. *N. Engl. J. Med.* 355, 560–569.
- Feinberg, A. P., and Tycko, B. (2004). The history of cancer epigenetics. *Nat. Rev. Cancer* 4, 143–153.
- Feng, P., and Navaratna, M. (2007). Modelling periodic oscillations during somitogenesis. *Math. Biosci. Eng.* 4, 661–673.
- Fu, G. F., Luo, J., Berg, A., Wang, Z., Li, J. H., Das, K., Li, R., and Wu, R. L. (2010). A dynamic model for functional mapping of biological rhythms. *J. Biol. Dyn.* 4, 1–10.
- Fu, G. F., Wang, Z., Li, J. H., and Wu, R. L. (2011). A mathematical framework for functional mapping of complex phenotypes using delay differential equations. *J. Theor. Biol.* 289, 206–216.
- Ganguly, R., and Puri, I. K. (2006). Mathematical model for the cancer stem cell hypothesis. *Cell Prolif.* 39, 3–14.
- Ganly, I., Talbot, S., Maghami, E., Viale, A., Pfister, D., Olshen, A. B., Shah, J. P., Succi, N., and Singh, B. (2007). Identification of angiogenesis/metastases genes predicting chemoradiotherapy response in patients with laryngopharyngeal carcinoma. *J. Clin. Oncol.* 25, 1369–1376.
- Goldbeter, A. (2002). Computational approaches to cellular rhythms. *Nature* 420, 238–245.
- Greenman, C., Stephens, P., Smith, R., Dalgliesh, G. L., Hunter, C., Bignell, G., Davies, H., Teague, J., Butler, A., Stevens, C., Edkins, S., O'Meara, S., Vastrik, I., Schmidt, E. E., Avis, T., Barthorpe, S., Bhamra, G., Buck, G., Choudhury, B., Clements, J., Cole, J., Dicks, E., Forbes, S., Gray, K., Halliday, K., Harrison, R., Hills, K., Hinton, J., Jenkinson, A., Jones, D., Menzies, A., Mironenko, T., Perry, J., Raine, K., Richardson, D., Shepherd, R., Small, A., Tofts, C., Varian, J., Webb, T., West, S., Widaa, S., Yates, A., Cahill, D. P., Louis, D. N., Goldstraw, P., Nicholson, A. G., Brasseur, F., Looijenga, L., Weber, B. L., Chiew, Y. E., DeFazio, A., Greaves, M. F., Green, A. R., Campbell, P., Birney, E., Easton, D. F., Chenevix-Trench, G., Tan, M. H., Khoo, S. K., Teh, B. T., Yuen, S. T., Leung, S. Y., Wooster, R., Futreal, P. A., and Stratton, M. R. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.
- Hartman, J. L., Garvik, B., and Hartwell, L. (2001). Principles for the buffering of genetic variation. *Science* 291, 1001–1004.
- Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* 4, 682–690.
- Jallepalli, P. V., and Lengauer, C. (2001). Chromosome segregation and cancer: Cutting through the mystery. *Nat. Rev. Cancer* 1, 109–117.
- Jirtle, R. L., and Skinner, M. K. (2007). Environmental epigenomics and disease susceptibility. *Nat. Rev. Genet.* 8, 253–262.
- Kaiser, J. (2005). Tackling the cancer genome. *Science* 309, 6.
- Khalil, I. G., and Hill, C. (2005). Systems biology for cancer. *Curr. Opin. Oncol.* 17, 44–48.
- Khan, J., Wei, J. S., Ringnr, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* 7, 673–679.
- Kirk, P. D. W., Toni, T., and Stumpf, M. P. H. (2008). Parameter inference for biochemical systems that undergo a Hopf bifurcation. *Biophys. J.* 95, 540–549.
- Lander, E. S., and Botstein, D. (1989). Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121, 185–199.
- Lapidot, T., Sirard, C., Vormoor, J., Murdoch, B., Hoang, T., Caceres-Cortes, J., Minden, M., Paterson, B., Caligiuri, M. A., and Dick, J. E. (1994). A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* 17, 645–648.
- Lehner, B., Crombie, C., Tischler, J., Fortunato, A., and Fraser, A. G. (2006). Systematic mapping of genetic interactions in *Caenorhabditis elegans* identifies common modifiers of diverse signaling pathways. *Nat. Genet.* 38, 896–903.
- Li, J. H., Das, K., Fu, G. F., Li, R., and Wu, R. L. (2011a). Bayesian lasso for genome-wide association studies. *Bioinformatics* 27, 516–523.
- Li, Y., Berg, A., Wu, L. R., Chen, G., and Wu, R. L. (2010a). Modeling the aneuploidy control of cancer. *BMC Cancer* 10, 346. doi:10.1186/1471-2407-10-346
- Li, Y., Guo, Y. Q., Hou, W., Chang, M., Liao, D. P., and Wu, R. L. (2011b). A statistical design for testing transgenerational genomic imprinting in natural human populations. *PLoS ONE* 6, e16858. doi:10.1371/journal.pone.0016858
- Li, Y., Ma, C. M., Wang, Z., Chen, G., Ahn, A., Lazarus, P., and Wu, R. L. (2010b). A disequilibrium model for detecting somatic mutations causing cancer. *J. Theor. Biol.* 165, 218–224.
- Li, Y., and Wu, R. L. (2009). Modeling host-cancer genetic interactions with multilocus sequence data. *J. Comput. Sci. Syst. Biol.* 2, 24–43.
- Li, Y., and Wu, R. L. (2010). Functional mapping of growth and development. *Biol. Rev.* 85, 207–216.
- Lin, D. Y., and Huang, B. E. (2007). The use of inferred haplotypes in downstream analyses. *Am. J. Hum. Genet.* 80, 577–579.
- Liu, T., Johnson, J. A., Casella, G., and Wu, R. L. (2004). Sequencing complex diseases with HapMap. *Genetics* 168, 503–511.
- Liu, T., Liu, X. L., Chen, Y. M., and Wu, R. L. (2007). A unifying differential equation model for functional genetic mapping of circadian rhythms. *Theor. Biol. Med. Model.* 4, 5.
- Lobo, N. A., Shimono, Y., Qian, D., and Clarke, M. F. (2007). The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol.* 23, 675–699.
- Lopez-Rios, F., Chuai, S., Flores, R., Shimizu, S., Ohno, T., Wakahara, K., Illei, P. B., Hussain, S., Krug, L., Zakowski, M. F., Rusch, V., Olshen, A. B., and Ladanyi, M. (2006). Global gene expression profiling of pleural mesotheliomas: overexpression of aurora kinases and P16/CDKN2A deletion as prognostic factors and critical evaluation of microarray-based prognostic prediction. *Cancer Res.* 66, 2970–2979.
- Luo, J. T., Hager, W. W., and Wu, R. L. (2010). A differential equation model for functional mapping of a virus-cell dynamic system. *J. Math. Biol.* 65, 1–15.
- Michor, F. (2008). Mathematical models of cancer stem cells. *J. Clin. Oncol.* 26, 2854–2860.
- Pearson, T. A., and Manolio, T. A. (2008). How to interpret a genome-wide association study. *J. Am. Med. Assoc.* 299, 1335–1344.
- Piotrowska, M. J., Enderling, H., an der Heiden, U., and Mackey, M. C. (2008). “Mathematical modeling of stem cells related to cancer,” in *Cancer and Stem Cells*, eds T. Dittmar, K. S. Zanker (New York: Nova Science Publishers), 11–35.
- Pujana, M. A., Han, J. D., Starita, L. M., Stevens, K. N., Tewari, M., Ahn JS, Rennert, G., Moreno, V., Kirchhoff, T., Gold, B., Assmann, V., Elshamy, W. M., Rual, J. F., Levine, D., Rozek, L. S., Gelman, R. S., Gunsalus, K. C., Greenberg, R. A., Sobhian, B., Bertin, N., Venkatesan, K., Ayivi-Guedehoussou, N., Solé, X., Hernández, P., Lázaro, C., Nathanson, K. L., Weber, B. L., Cusick, M. E., Hill, D. E., Offit, K., Livingston, D. M.,

- Gruber, S. B., Parvin, J. D., and Vidal, M. (2007). Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat. Genet.* 39, 1338–1349.
- Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. G. (2007). Parameter estimation for differential equations: a generalized smoothing approach (with discussion). *J. R. Stat. Soc. B* 69, 741–796.
- Reik, W., and Lewis, A. (2005). Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat. Rev. Genet.* 6, 403–410.
- Rhodes, D. R., and Chinnaiyan, A. M. (2005). Integrative analysis of the cancer transcriptome. *Nat. Genet.* 37: S31–S37.
- Rual, J. F., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li N Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albalá, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M. E., Hill, D. E., Roth, F. P., and Vidal, M. (2005). Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 437, 1173–1178.
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* 2008, 129–140.
- Savcenko, V. (2009). Construction of a multirate RODAS method for stiff ODEs. *J. Comput. Appl. Math.* 225, 323–337.
- Secko, D. (2005). How an imprint can lead to cancer. *Can. Med. Assoc. J.* 172, 1286.
- Sha, K. (2008). A mechanistic view of genomic imprinting. *Annu. Rev. Genomics Hum. Genet.* 9, 197–216.
- Singh, S. K., Hawkins, C., Clarke, I. D., Squire, J. A., Bayani, J., Hide, T., Henkelman, R. M., Cusimano, M. D., and Dirks, P. B. (2004). Identification of human brain tumour initiating cells. *Nature* 432, 396–401.
- Solyani, G. I., Berezetskaya, N. M., Bulkiewicz, R. I., and Kulik, G. I. (1995). Different growth patterns of a cancer cell population as a function of its starting growth characteristics. Analysis by mathematical modelling. *Cell Prolif.* 28, 263–278.
- Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F. H., Goehler H Stroedicke, M., Zenkner, M., Schoenherr, A., Koepfen, S., Timm, J., Mintzlaff, S., Abraham, C., Bock, N., Kietzmann, S., Goedde, A., Toksoez, E., Droege, A., Krobitsch, S., Korn, B., Birchmeier, W., Lehrach, H., and Wanker, E. E. (2005). A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 122, 957–968.
- Stock, R. P., and Bialy, H. (2002). The sigmoidal curve of cancer. *Nat. Biotechnol.* 21, 13–14.
- The International HapMap Consortium. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449, 851–862.
- Verdugo, A., and Rand, R. (2008). Hopf bifurcation in a DDE model of gene expression. *Commun. Nonlinear Sci. Numer. Simul.* 13, 235–242.
- Viale, A., De Franco, F., Orleth, A., Cambiagli, V., Giuliani, V., Bossi, D., Ronchini, C., Ronzoni, S., Muradore, I., Monestiroli, S., Gobbi, A., Alcalay, M., Minucci, S., and Pelicci, P. G. (2009). Cell-cycle restriction limits DNA damage and maintains self-renewal of leukaemia stem cells. *Nature* 457, 51–56.
- Visvader, J. E., and Lindeman, G. J. (2008). Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat. Rev. Cancer* 8, 755–768.
- Wilkins, J. F., and Haig, D. (2003). What good is genomic imprinting: The function of parent-specific gene expression. *Nat. Rev. Genet.* 4, 359–368.
- Wu, R., Cao, J., Huang, Z., Wang, Z., Gai, J., and Vallejos, E. (2011). Systems mapping: how to improve the genetic mapping of complex traits through design principles of biological systems. *BMC sys. biol.* 5, 84–94. doi: 10.1186/1752-0509-5-84
- Wu, R. L., and Lin, M. (2006). Functional mapping – how to map and study the genetic architecture of dynamic complex traits. *Nat. Rev. Genet.* 7, 229–237.
- Wu, R. L., Ma, C. X., Lin, M., and Casella, G. (2004). A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* 166, 1541–1551.
- Yap, J. S., Fan, J., and Wu, R. L. (2009). Nonparametric modeling of covariance structure in functional mapping of quantitative trait loci. *Biometrics* 65, 1068–1077.
- Youngson, N. A., and Whitelaw, E. (2008). Transgenerational epigenetic effects. *Annu. Rev. Genomics Hum. Genet.* 9, 233–257.
- Zhao, W., Chen, Y. Q., Casella, G., Cheverud, J. M., and Wu, R. L. (2005). A nonstationary model for functional mapping of complex traits. *Bioinformatics* 21, 2469–2477.
- Zimmerman, D. L., and Nunez-Anton, V. (2001). Parametric modeling of growth curve data: An overview (with discussion). *Test* 10, 1–73.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 11 January 2012; accepted: 26 April 2012; published online: 22 May 2012.

Citation: Wang Z, Liu J, Wang J, Wang Y, Wang N, Li Y, Li R and Wu R (2012) Dynamic modeling of genes controlling cancer stem cell proliferation. *Front. Genet.* 3:84. doi: 10.3389/fgene.2012.00084

This article was submitted to *Frontiers in Statistical Genetics and Methodology*, a specialty of *Frontiers in Genetics*. Copyright © 2012 Wang, Liu, Wang, Wang, Wang, Li, Li and Wu. This is an open-access article distributed under the terms of the Creative Commons Attribution Non Commercial License, which permits non-commercial use, distribution, and reproduction in other forums, provided the original authors and source are credited.